

PATENT
Attorney Docket No.: 16869W-
110400US
Client Ref. No.: P04032USA

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

MISAKO TAMURA et al.

Application No.: 10/803,204

Filed: March 17, 2004

For: INFORMATION PROCESSING
DEVICE AND METHOD

Customer No.: 20350

Examiner: Unassigned

Technology Center/Art Unit: 2182

Confirmation No.: 5843

**PETITION TO MAKE SPECIAL FOR
NEW APPLICATION UNDER M.P.E.P.
§ 708.02, VIII & 37 C.F.R. § 1.102(d)**

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is a petition to make special the above-identified application under MPEP § 708.02, VIII & 37 C.F.R. § 1.102(d). The application has not received any examination by an Examiner.

(a) The Commissioner is authorized to charge the petition fee of \$130 under 37 C.F.R. § 1.17(i) and any other fees associated with this paper to Deposit Account 20-1430.

03/07/2005 BABRAHA1 00000079 201430 10803204

01 FC:1464 130.00 DA

Effective on 12/08/2004.

Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL

For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT (\$) **130.00**
Complete if Known

Application Number	10/803,204
Filing Date	March 17, 2004
First Named Inventor	Tamura, Misako
Examiner Name	Unassigned
Art Unit	2182
Attorney Docket No.	16869W-110400US

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____
☒ Deposit Account Deposit Account Number: 20-1430 Deposit Account Name: Townsend and Townsend and Crew LLP

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

☒ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, except for the filing fee

☒ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17 ☒ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038
FEE CALCULATION
1. BASIC FILING, SEARCH, AND EXAMINATION FEES

Application Type	FILING FEES		SEARCH FEES		EXAMINATION FEES		Fees Paid (\$)
	Small Entity	Fee (\$)	Small Entity	Fee (\$)	Small Entity	Fee (\$)	
Utility	300	150	500	250	200	100	
Design	200	100	100	50	130	65	
Plant	200	100	300	150	160	80	
Reissue	300	150	500	250	600	300	
Provisional	200	100	0	0	0	0	

2. EXCESS CLAIM FEES

Fee Description	Small Entity	
	Fee (\$)	Fee (\$)
Each claim over 20 or, for Reissues, each claim over 20 and more than in the original patent	50	25
Each independent claim over 3 or, for Reissues, each independent claim more than in the original patent	200	100
Multiple dependent claims	360	180

Total Claims	Extra Claims	Fee (\$)	Fee Paid (\$)	Multiple Dependent Claims	Fee (\$)	Fee Paid (\$)
-20 or HP = _____ x _____ = _____						
HP = highest number of total claims paid for, if greater than 20						
Indep. Claims	Extra Claims	Fee (\$)	Fee Paid (\$)			
-3 or HP = _____ x _____ = _____						
HP = highest number of independent claims paid for, if greater than 3						

3. APPLICATION SIZE FEE

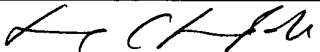
If the specification and drawings exceed 100 sheets of paper, the application size fee due is \$250 (\$125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	Extra Sheets	Number of each additional 50 or fraction thereof	Fee (\$)	Fee Paid (\$)
_____, - 100 = _____ / 50 = _____ (round up to a whole number) x _____ = _____				

4. OTHER FEE(S)

Non-English Specification, \$130 fee (no small entity discount)

 Other: PETITIONS TO THE COMMISSIONER
Fees Paid (\$)
130.00
SUBMITTED BY

Signature		Registration No. (Attorney/Agent) 41,405	Telephone 650-326-2400
Name (Print/Type)	Chun-Pok Leung		Date February 28, 2005

(b) All the claims are believed to be directed to a single invention. If the Office determines that all the claims presented are not obviously directed to a single invention, then Applicants will make an election without traverse as a prerequisite to the grant of special status.

(c) Pre-examination searches were made of U.S. issued patents, including a classification search, a computer database search, and a keyword search. The searches were performed on or around January 27, 2005, and were conducted by a professional search firm, Kramer & Amado, P.C. The classification search covered Class 711 (subclasses 112 and 113) for the U.S. and foreign subclasses identified above. The computer database search was conducted on the USPTO systems EAST and WEST, as well as through commercial databases for non-patent literature. The keyword search was conducted in Class 710 (subclasses 5, 17, 36, 40, 45, and 126) and Class 711 (subclasses 114, 119, 138, 141, 147, 162, and 202). The inventors further provided a reference considered most closely related to the subject matter of the present application (see reference #5 below), which was cited in the Information Disclosure Statements filed on March 17, 2004.

(d) The following references, copies of which are attached herewith, are deemed most closely related to the subject matter encompassed by the claims:

- (1) U.S. Patent No. 6,298,418 B1;
- (2) U.S. Patent Publication No. 2001/0034801 A1;
- (3) U.S. Patent No. 6,715,006 B1;
- (4) U.S. Patent No. 5,761,515; and
- (5) Japanese Patent Publication No. JP 2002-140233.

(e) Set forth below is a detailed discussion of references which points out with particularity how the claimed subject matter is distinguishable over the references.

A. Claimed Embodiments of the Present Invention

The claimed embodiments relate to technology for processing an information set having one or more information elements, and more specifically, to technology for

processing an I/O request received from an external device in accordance with a protocol such as an ESCON or FICON, or the like, for example.

Independent claim 1 recites an information processing device comprising a receiving component receiving information elements contained in respective information sets having one or more information elements, from one or a plurality of information set sources issuing the information sets; an information processing component carrying out processing of the information elements thus received; and a determining component determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the length of processing time for the two or more information sets becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets to be processed in accordance with the reception sequence thereof. The information processing component starts processing of the plurality of information elements that are unprocessed or currently being processed, on the basis of the processing sequence thus determined.

Independent claim 12 recites a storage control apparatus comprising a storage control device capable of connecting to one or a plurality of physical or logical storage devices; a receiving component receiving information elements contained in respective I/O requests having one or more information elements, from one or a plurality of I/O request sources issuing the I/O requests; an information processing component carrying out processing for reading data from the storage device and transmitting same to the I/O request source, or writing data from the I/O request source to the storage device, on the basis of the information elements contained in the I/O request thus received; and a determining component determining a processing sequence for the two or more I/O requests or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more I/O requests thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more I/O requests becomes equal to or less than the value that would be obtained were the plurality of

information elements or the two or more I/O requests to be processed according to the reception sequence thereof.

Independent claim 13 recites an information processing method comprising receiving information elements contained in respective information sets having one or more information elements, from one or a plurality of information set sources issuing the information sets; carrying out processing of the information elements thus received; and determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets to be processed in accordance with the reception sequence thereof. In the processing of the information elements, the processing of a plurality of information elements that are unprocessed or currently being processed is started on the basis of the element processing sequence thus determined.

Independent claim 14 recites a computer-readable storage medium having a computer program comprising code for receiving information elements contained in respective information sets having one or more information elements, from one or a plurality of information set sources issuing the information sets; code for carrying out processing of the information elements thus received; and code for determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets to be processed in accordance with the reception sequence thereof. In the processing of the information elements, the processing of a plurality of information elements that are unprocessed or currently being processed is started on the basis of the element processing sequence thus determined.

One of the benefits that may be derived is that it is possible to reduce the value relating to the average of the processing times taken to process a plurality of information sets.

B. Discussion of the References

1. U.S. Patent No. 6,298,418 B1

This reference relates to a memory write control method for a multiprocessor system including a plurality of processor modules sharing at least one memory module via a crossbar switch, the memory write control method comprising the steps of: upon the one of the plurality of processor modules which issued the memory write request receiving a notification of the completion of the memory write request from a crossbar switch, the one of the plurality of processor modules, as a response to the cache coherency check request, notifying the any one of the plurality of processor modules which issued the cache coherency check request that updated data in a cache memory is invalid if the updated data at the address corresponding to the cache miss is latest updated data exclusively possessed by the one of the plurality of processor modules which issued the memory write request and if the memory write request was for the updated data.

The reference does not teach determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets (or I/O requests) thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets (or I/O requests) becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets (or I/O requests) to be processed in accordance with the reception sequence thereof, as recited in independent claims 1, 12, 13, and 14.

2. U.S. Patent Publication No. 2001/0034801 A1

This reference discloses an apparatus for processing an input/output request from an upper unit by using a plurality of channel buses comprising, a busy ratio measurement storing section that calculates the busy ratio by dividing the measured number of busy response times by the number of input/output activation times from the channel unit

every input/output port and stores; and route control section selects the maximum time when the busy ratio of the low speed input/output port is larger than the busy ratio of the high speed input/output port, selects the average time when the busy ratio of the low speed input/output port is equal to the busy ratio of the high speed input/output port, and selects the minimum time when the busy ratio of the low speed input/output port is smaller than the busy ratio of the high speed input/output port. See figures and paragraphs [0008]-[0013].

The reference does not teach determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets (or I/O requests) thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets (or I/O requests) becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets (or I/O requests) to be processed in accordance with the reception sequence thereof, as recited in independent claims 1, 12, 13, and 14.

3. U.S. Patent No. 6,715,006 B1

This reference discloses a disk time-sharing apparatus and method. A disk apparatus includes a re-ordering function for rearranging execution waiting inputs and outputs so as to minimize the processing time. The re-ordering function is such a function that an input/output to minimize a positioning time that is defined by the sum of a seeking time and a rotation waiting time is selected by the disk apparatus as an input/output to be executed next from the execution waiting inputs/outputs. When the input/output is requested to the disk apparatus, a simple task serving as a task designation indicating that the input/output can be set as a target of the re-ordering is notified to the disk apparatus. In case of the inputs/outputs of the simple task designation, the disk apparatus schedules the inputs and outputs so as to minimize the positioning time. The average processing time at the time of the random access is reduced.

The reference does not teach determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or

more information sets (or I/O requests) thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets (or I/O requests) becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets (or I/O requests) to be processed in accordance with the reception sequence thereof, as recited in independent claims 1, 12, 13, and 14.

4. U.S. Patent No. 5,761,515

This reference discloses a computer processing unit for tolerating cache miss latency by dynamically switching appropriately between two different code sequences, one optimized at compile-time, assuming a cache-hit, and the other optimized at compile-time, assuming a cache-miss.

The reference does not teach determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets (or I/O requests) thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets (or I/O requests) becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets (or I/O requests) to be processed in accordance with the reception sequence thereof, as recited in independent claims 1, 12, 13, and 14.

5. Japanese Patent Publication No. JP 2002-140233

This reference relates to a technique to suppress the increase of response time due to an execution of staging following a cache miss. In the information processing system, the storage subsystem composed of a disk controller 2 provided with a cache memory 24 and a subordinate disk device 3 is connected to a central processor 1 by an interface such as an FC-SB2 where the central processor 1 issues an I/O request composed of a plurality of commands and a chain of data asynchronously to a response from the disk controller 2. The disk controller 2 is provided with a function concurrently carrying out a process executing a command where object data hits the cache memory 24 and a process staging object data of a command of a cache miss from the disk device 3 to the cache memory 24.

As discussed in the present application at page 3, line 22 to page 4, line 10, the reference discloses technology aimed at reducing the I/O response time. According to this technology, for example, of a plurality of I/O requests received from an host, the storage control device carries out data transfer to the host in respect of those I/O requests which produce "hits" in the cache, and executes parallel processing for reading corresponding data from the data storage device to the cache memory, and then transferring this data to the host, in respect of I/O requests which produce "misses" in the cache. As a result, the completion notification sequence sent by the memory control device to the host in respect of processing a plurality of I/O requests, may be different from the sequence in which the I/O requests were received from the host.

The reference does not teach determining a processing sequence for the two or more information sets or the plurality of information elements, on the basis of the plurality of information elements that are unprocessed or currently being processed, contained in two or more information sets (or I/O requests) thus received, and determining a processing sequence different from a reception sequence, in which a value relating to the average of the response time for the two or more information sets (or I/O requests) becomes equal to or less than the value that would be obtained were the plurality of information elements or the two or more information sets (or I/O requests) to be processed in accordance with the reception sequence thereof, as recited in independent claims 1, 12, 13, and 14.

(f) In view of this petition, the Examiner is respectfully requested to issue a first Office Action at an early date.

Respectfully submitted,



Chun-Pok Leung
Reg. No. 41,405

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400
Fax: 415-576-0300
Attachments
RL:rl
60423348 v1

P04032

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-140233

(43)Date of publication of application : 17.05.2002

(51)Int.Cl.

G06F 12/08
G06F 3/06

(21)Application number : 2000-332164

(71)Applicant : HITACHI LTD

(22)Date of filing : 31.10.2000

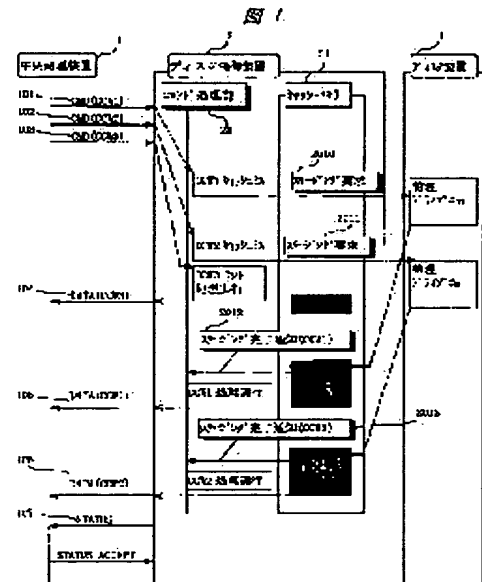
(72)Inventor : FURUUMI NOBORU
AZUMI YOSHIHIRO

(54) STORAGE SUB SYSTEM, CONTROL METHOD FOR I/O INTERFACE AND INFORMATION PROCESSING SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To suppress the increase of response time due to an execution of staging following a cache miss.

SOLUTION: In the information processing system, the storage sub system composed of a disk controller 2 provided with a cache memory 24 and a subordinate disk device 3 is connected to a central processor 1 by an interface such as an FC-SB2 where the central processor 1 issues an I/O request composed of a plurality of commands and a chain of data asynchronously to a response from the disk controller 2. The disk controller 2 is provided with a function concurrently carrying out a process executing a command where object data hits the cache memory 24 and a process staging object data of a command of a cache miss from the disk device 3 to the cache memory 24.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-140233
(P2002-140233A)

(43) 公開日 平成14年5月17日 (2002.5.17)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード (参考)	
G 0 6 F 12/08	5 5 7	G 0 6 F 12/08	5 5 7	5 B 0 0 5
	5 1 9		5 1 9 B	5 B 0 6 5
			5 1 9 Z	
	5 4 1		5 4 1 Z	
3/06	3 0 2	3/06	3 0 2 A	
審査請求 未請求 請求項の数10 O L (全 20 頁) 最終頁に続く				

(21) 出願番号 特願2000-332164(P2000-332164)

(22) 出願日 平成12年10月31日 (2000. 10. 31)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 古海 昇

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 安積 義弘

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 100080001

弁理士 筒井 大和

Fターム(参考) 5B005 JJ11 KK03 KK15 MM11 SS12

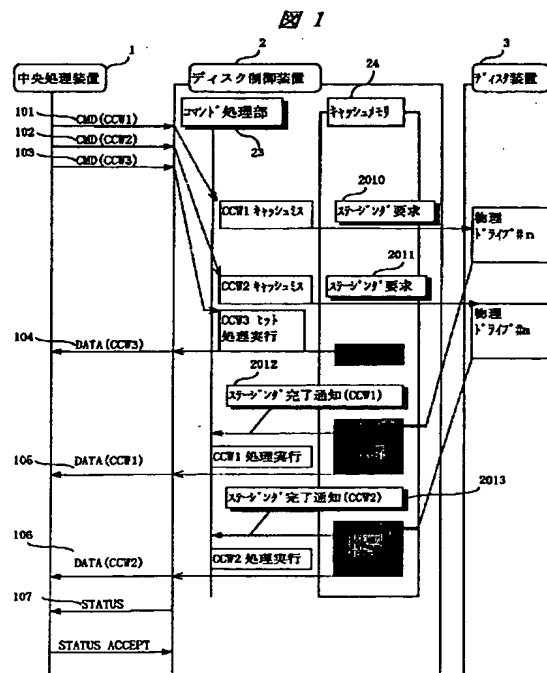
5B065 BA01 CH01

(54) 【発明の名称】 記憶サブシステム及びI/Oインタフェースの制御方法ならびに情報処理システム

(57) 【要約】

【課題】 キャッシュミスに伴うステージング処理実行によるレスポンスタイムの増加を抑えること。

【解決手段】 キャッシュメモリ24を備えたディスク制御装置2および配下のディスク装置3から構成される記憶サブシステムを、当該ディスク制御装置2からの応答とは非同期に、中央処理装置1が複数のコマンド及びデータのチェーンで構成されるI/O要求を発行するFC-SB2等のインタフェースで当該中央処理装置1に接続する情報処理システムにおいて、ディスク制御装置2は、中央処理装置1からの複数のコマンドの受け付け順に関係なく、対象データがキャッシュメモリ24にヒットしたコマンドを実行する処理と、キャッシュミスのコマンドの対象データをディスク装置3からキャッシュメモリ24にステージングする処理とを並行して行う機能を備えた。



【特許請求の範囲】

【請求項 1】 中央処理装置に接続され、配下に複数の記憶装置を有し、内部に前記中央処理装置と前記記憶装置との間で授受されるデータが一時的に格納されるキャッシュメモリを有する記憶制御装置を含み、前記記憶制御装置からの応答とは非同期に、前記中央処理装置が複数のコマンド及びデータのチェーンで構成される I/O 要求を前記記憶制御装置に対して発行する I/O インタフェースプロトコルにより、前記中央処理装置に接続されている記憶サブシステムであって、

前記記憶制御装置は、前記中央処理装置からの複数の前記コマンド及びデータの受領順には依存せずに、受領した前記コマンドの処理順序を決定して実行する手段を有することを特徴とする記憶サブシステム。

【請求項 2】 請求項 1 記載の記憶サブシステムにおいて、

前記記憶制御装置は、前記中央処理装置から受領した複数のコマンドのうち、処理対象のデータが前記キャッシュメモリ上に存在する前記コマンドについては、前記中央処理装置とのデータ転送を実行し、処理対象のデータが前記キャッシュメモリ上に存在しない前記コマンドの処理対象データを、前記記憶装置から前記キャッシュメモリへ読み出す処理を、前記中央処理装置とのデータ転送処理と並行して実行する手段を有することを特徴とする記憶サブシステム。

【請求項 3】 請求項 1 記載の記憶サブシステムにおいて、

前記記憶制御装置は、1つの論理ボリュームのデータを複数の前記記憶装置に分散して格納する RAID 構成をとっている場合において、前記中央処理装置から受領した複数のコマンドをまとめてコマンド群とし、前記コマンド群のうちの複数のコマンドの処理対象データが前記キャッシュメモリ上に存在しない場合、前記各コマンドの処理対象データが格納されている個々の前記記憶装置に対して、前記コマンド群の各コマンドの処理対象データの読み出し処理を複数並行して起動し、前記読み出し処理が完了した順に前記コマンド群の各コマンドに関する前記中央処理装置とのデータ転送と前記読み出し処理を並行して実行する手段を有することを特徴とする記憶サブシステム。

【請求項 4】 請求項 3 記載の記憶サブシステムにおいて、

前記記憶制御装置は、前記中央処理装置から受領した複数のコマンドをまとめてコマンド群とし、前記コマンド群のうちの複数のコマンドの処理対象データが前記キャッシュメモリ上に存在しない場合、前記コマンドの各々の処理対象データが格納されている前記記憶装置の各々の稼働率に応じて前記記憶装置から前記キャッシュメモリへの前記処理対象データの読み出し処理を複数並行して起動し、前記読み出し処理が完了したものから、前記中央

処理装置とのデータ転送を他の前記読み出し処理と並行して実行する手段を有することを特徴とする記憶サブシステム。

【請求項 5】 請求項 4 記載の記憶サブシステムにおいて、

前記記憶制御装置は、前記中央処理装置から受領した複数のコマンドを纏めて第 1 のコマンド群とし、前記第 1 のコマンド群のうちの複数のコマンドの処理対象データが前記キャッシュメモリ上に存在しない場合、前記第 1 のコマンド群のうち、処理対象データが同一の前記記憶装置に存在しているコマンドを前記記憶装置へアクセスするアドレスの昇順にまとめて第 2 のコマンド群とし、前記第 2 のコマンド群単位で前記記憶装置から前記キャッシュメモリへの前記記憶装置内のデータの読み出し範囲を決定して前記読み出し処理を起動し、前記読み出し処理が完了したものから前記中央処理装置へのデータ転送を他の前記読み出し処理と並行して実行する手段を有することを特徴とする記憶サブシステム。

【請求項 6】 中央処理装置と、前記中央処理装置との間で授受されるデータが格納される複数の記憶装置を配下にもち、前記データが一時的に格納されるキャッシュメモリを備えた記憶制御装置との接続に用いられ、前記記憶制御装置からの応答とは非同期に、前記中央処理装置が複数のコマンド及びデータのチェーンで構成される I/O 要求を前記記憶制御装置に対して発行するプロトコルを備えた I/O インタフェースの制御方法であって、

前記記憶制御装置は、前記中央処理装置からのコマンドに関して前記コマンドの受領順に依存しない順序で前記中央処理装置とデータ転送を実行し前記コマンド毎にコマンド終了報告を行ない、

前記中央処理装置は、当該中央処理装置が発行したコマンド及びデータの発行順には依存しない順で、前記記憶制御装置からのデータ及びコマンド終了報告を受領し、前記受領したデータ及びコマンド終了報告に対応する発行済みコマンドを特定し、前記データ及びコマンド終了報告を前記特定したコマンドに対する応答フレームとして処理すること、を特徴とする I/O インタフェースの制御方法。

【請求項 7】 請求項 6 記載の I/O インタフェースの制御方法において、

前記記憶制御装置は、前記中央処理装置から受領したコマンド及びデータのチェーンの途中でエラー及びリトライ要因が発生した場合、前記 I/O 要求を中断せずに継続して受領済みの他の実行可能な前記コマンド及びデータに関して前記中央処理装置とのデータ転送を実施し、前記 I/O 要求を構成する全コマンド処理分の終了状態を 1 つにまとめて I/O 要求処理完了報告として前記中央処理装置へ報告し、

前記中央処理装置は、前記 I/O 要求処理完了報告を受

領後、前記各コマンドの終了状態を認識し、エラーあるいはリトライ要求のあった前記各コマンドに対してのみリカバリ処理を実行すること、を特徴とするI/Oインタフェースの制御方法。

【請求項8】 請求項6記載のI/Oインタフェースの制御方法において、

前記中央処理装置は、当該中央処理装置が発行したコマンド及びデータの発行順には依存しない順で、前記記憶制御装置からデータ及びコマンド終了報告を受領する処理と、前記記憶制御装置から前記I/O要求を構成する全コマンド分に対するコマンド終了報告を受領するまで、前記I/O要求処理を中断せず継続して実行する処理と、前記I/O要求を構成する全コマンド分の前記コマンド終了報告を受領後、受領した前記コマンド終了報告の中でエラーあるいはリトライ要求があったコマンドに対してのみ、リカバリ処理を実行する処理と、を行うことを特徴とするI/Oインタフェースの制御方法。

【請求項9】 中央処理装置と、前記中央処理装置との間で授受されるデータが格納される複数の記憶装置を配下にもち、前記データが一時的に格納されるキャッシュメモリを備えた記憶制御装置と、前記中央処理装置と前記記憶制御装置とを接続し、前記記憶制御装置からの応答とは非同期に、前記中央処理装置が複数のコマンド及びデータのチェーンで構成されるI/O要求を前記記憶制御装置に対して発行するプロトコルを備えたI/Oインタフェースとを含む情報処理システムであって、前記記憶制御装置は、前記中央処理装置からのコマンドに関して前記コマンドの受領順に依存しない順序で前記中央処理装置とデータ転送を実行し前記コマンド毎にコマンド終了報告を行なう手段を備え、前記中央処理装置は、当該中央処理装置が発行したコマンド及びデータの発行順には依存しない順で、前記記憶制御装置からのデータ及びコマンド終了報告を受領し、前記受領したデータ及びコマンド終了報告に対応する発行済みコマンドを特定し、前記データ及びコマンド終了報告を前記特定したコマンドに対する応答フレームとして処理する手段を備えた、ことを特徴とする情報処理システム。

【請求項10】 請求項9記載の情報処理システムにおいて、

前記記憶制御装置は、前記中央処理装置から受領したコマンド及びデータのチェーンの途中でエラー及びリトライ要因が発生した場合、前記I/O要求を中断せずに継続して受領済みの他の実行可能な前記コマンド及びデータに関して前記中央処理装置とのデータ転送を実施し、前記I/O要求を構成する全コマンド処理分の終了状態を1つにまとめてI/O要求処理完了報告として前記中央処理装置へ報告する手段を備え、前記中央処理装置は、前記I/O要求処理完了報告を受領後、前記各コマンドの終了状態を認識し、エラーある

いはリトライ要求のあった前記各コマンドに対してのみリカバリ処理を実行する手段を備えた、ことを特徴とする情報処理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、記憶サブシステム及びI/Oインタフェースの制御技術ならびに情報処理システムに関し、より詳しくは、中央処理装置と記憶サブシステムの記憶制御装置とが、当該中央処理装置が複数のコマンドやデータから構成される1つのI/O要求を記憶制御装置からの応答とは非同期に記憶制御装置に対して発行することでI/O要求処理を実行するインタフェースプロトコルにより接続された情報処理システム等に適用して有効な技術に関する。

【0002】

【従来の技術】 銀行のオンラインシステムなどで用いられる大規模情報システム（メインフレームシステム）は、中央処理装置と周辺記憶装置とから構成される。この周辺記憶装置は、記憶制御装置と記憶装置から構成されており、これを記憶サブシステムと呼んでいる。以下、上記の記憶サブシステムとメインフレーム間のインタフェースについての概要を説明する。

【0003】 上記メインフレーム向け記憶サブシステムを構成する中央処理装置と記憶制御装置との間で、I/O要求処理の際に伝達される情報としては主に、（1）コマンド、（2）コマンド応答、（3）コマンド応答受け付け、（4）データ、（5）ステータスなどがあり、これらがフレームの形式で伝達されてI/O要求処理が実行される。

【0004】 中央処理装置は、記憶装置に対するI/O要求を実行するためにCCWチェーンと呼ばれる複数のコマンド及びデータから構成されるコマンド群を作成する。中央処理装置は、このコマンド群の最初のコマンドを記憶制御装置に対して発行する。これに対し、コマンドを受領した記憶制御装置は、コマンドフレームを受領したことを通知するコマンド応答フレームを中央処理装置に対して送信する。このコマンド応答フレームに対して、中央処理装置はコマンド応答受け付けフレームを記憶制御装置に対して送信する。この時点で、中央処理装置及び記憶制御装置は、共にデータの送受信が可能な状態になったことを認識し、この後、中央処理装置と記憶制御装置間でデータの送受信が開始される。発行されたコマンドに関するデータの送受信が終了した時点で、記憶制御装置から中央処理装置に対してデータ転送処理の終了状態を通知するステータスフレームが送信される。

【0005】 中央処理装置は、記憶制御装置からのステータスフレームを受領した後、このステータスの内容をチェックし、次コマンド処理が継続可能であるならば、次コマンドを発行する。この様に、1つのCCWチェーンは、中央処理装置と記憶制御装置間において、1つ1

つのコマンド毎にコマンド～コマンド応答～データ転送～ステータス送信といったインターロックをとりながら、逐次処理されていく。

【0006】ここで、もう少しCCWチェーンについて詳細に述べる。CCWチェーンを構成するコマンドの種類としては、レコードへのアクセス可否やアクセスモード等を指定するDefine Extentコマンド（以下、DXコマンドとする）や、対象となる入出力データのシリンダ・トラック・レコードに位置付けるための情報などを示すLocate Recordコマンド（以下、LOCコマンドとする）、そして実際のリード・ライトを指示するリード・ライトコマンド、等がある。

【0007】1つのCCWチェーンは、これらの複数コマンドのチェーンにより構成される。LOCコマンドを受領すると、記憶制御装置は、LOCコマンドのパラメータデータから位置付けすべきシリンダ・トラック・レコードを認識し、位置付け処理を行なう。

【0008】LOCコマンドの後続には、リード／ライトコマンドがチェーンする。LOCコマンドにチェーンしたリード／ライトコマンドの処理は、LOCコマンドで位置付けたレコードから連続したレコードに対して実行される。この様にLOCコマンドに続いてチェーンされたリード／ライトコマンドのコマンド群をLOCドメインという。LOCドメイン数、つまりLOCコマンドに幾つのリード／ライトコマンドがチェーンするかは、LOCコマンドのパラメータで指定される。

【0009】今、あるI/O要求を実行する1CCWチェーンにおいて、次に処理すべき対象レコードが、直前に処理したレコードとは不連続の場合、同一LOCドメインでは処理出来ず、次に処理すべきレコードへの位置付け処理が必要となる。この場合、再度LOCコマンドにて次に処理すべきレコードへの位置付けを行なう。この様に、1つのCCWチェーンの処理において、幾つかの不連続なレコードに対するリード・ライト要求がある場合には、1CCWチェーンに複数のLOCドメインが存在することになる。

【0010】次に、上記CCWチェーン実行時における、中央処理装置と記憶制御装置間の論理的な接続の切り離し動作について説明する。

【0011】中央処理装置から記憶制御装置配下のある記憶装置に対してリード／ライトコマンドが発行された時、記憶制御装置内のキャッシュメモリ上に処理対象のデータが存在しない場合には、記憶装置からデータをキャッシュメモリにステージングする必要がある。この場合には、記憶制御装置は前記のコマンド処理を直ぐには実行出来ない。このため、記憶制御装置は、一旦中央処理装置と記憶制御装置間の論理的な接続を切断する事を要求するステータスを中央処理装置に対して送信し、論理的な接続を切断する。その後、記憶制御装置内キャッ

シュメモリへのステージング処理が完了し、I/Oを処理する準備が出来た時点で、記憶制御装置は中央処理装置に対し接続割り込み要求の送信を行って論理的な接続を行った後、I/O処理再開を意味する状態通知を行なう。

【0012】この様に記憶制御装置は、I/O処理のための準備が出来ていない等の理由で一旦、中央処理装置との論理的な接続を切断するケースがある。この様な切り離しの要因としては、（1）記憶制御装置内のキャッシュメモリ上にデータが存在せず、記憶装置からデータを記憶制御装置内のキャッシュメモリにステージングするケース、（2）記憶制御装置内のキャッシュメモリのスペース割り当てが出来ず、キャッシュメモリスペースの空き待ちのケース、（3）I/O処理のための資源がBusy状態で確保出来ず、資源のBusy解除待ちのケース、等がある。

【0013】1CCWチェーンの実行中に、この様な切り離し動作が多く発生すると、I/O要求処理のトータルレスポンスタイムが増加してしまう。

【0014】次にキャッシュミスによるレスポンスタイムの増加を削減する技術の一例について述べる。

【0015】I/O要求のパターンの1つとして、大容量バッチ処理等に代表される様な、記憶装置内のレコードに対して、シーケンシャルにアクセスして処理するパターンがある。この場合のCCWチェーンは、前述のDXコマンド、LOCコマンド、そしてこのLOCコマンドにチェーンした複数のリードorライトコマンドから構成されており、連続したレコード・トラックの処理を実行するという特徴がある。処理対象のレコードは連続しているため、LOCドメインを切り替える必要はなく、連続してリード／ライトコマンドの処理が実行可能である。

【0016】先程述べた様に、リード／ライト対象レコードがキャッシュミスの場合は、中央処理装置と記憶制御装置との論理的な切り離しが発生するため、レスポンスタイムが増加してしまう。しかし、このシーケンシャルアクセスの場合は、CCWチェーンの次のコマンドを受領していなくても次にアクセスするシリンダ・トラック・レコードを予測出来るため、処理を行なうであろうシリンダ・トラック・レコードのデータを、前もってキャッシュ上にステージングしておくことにより、キャッシュミスによる中央処理装置と記憶制御装置との切り離しを実行する契機を削減でき、レスポンスタイムの向上が望める。ちなみに、当該CCWチェーンがシーケンシャルアクセスか否かは、DXコマンドにシーケンシャルアクセスを示す情報があるので、これを参照すればよい。

【0017】一方、別のI/O要求のパターンの1つとして、データベースへのアクセスに代表される様な、ランダムなレコードへのアクセスがある。ランダムアクセ

スでは、アクセス対象のレコードが分散しているため、それぞれのレコードの処理を行なう前に、L O Cコマンドにて位置付け処理を行なう必要がある。このため、ランダムアクセスのC C Wチェーンの中には、複数のL O Cドメインが存在する。シーケンシャルアクセスと異なり、ランダムアクセスでは、処理対象のレコードが連続していないために次にアクセスするレコードを予測出来ず、シーケンシャルアクセス時の様にアクセス対象のデータを先にステージングすることが出来ない。従って、ランダムアクセスの方がシーケンシャルアクセスより、キャッシュミスによる中央処理装置との論理的な切り離し契機が多くなる可能性があると言える。

【0018】以上、キャッシュミスによるレスポンスタイムの増加を軽減する技術について述べてきたが、次にスループット向上に関する技術について述べる。

【0019】近年の大容量転送、遠隔データ転送などを実現するプロトコルとして注目を浴びているのが、ファイバーチャネルプロトコルである。ファイバーチャネルプロトコルは、主に、これまでオープン系システムで使われてきた技術だが、最近、メインフレーム用ファイバーチャネルプロトコルとして、ファイバーチャネルプロトコルの物理層(F C-P H)に準拠したプロトコルであるF C-S B 2(F I B R E C H A N N E L S i n g l e-B y t e C o m m a n d C o d e S e t s-2 M a p p i n g P r o t o c o l)が提案されている。これは、F C-P Hに従来のメインフレームと記憶サブシステム間の通信プロトコルをマッピングしたものであり、現在、A N S I(A m e r i c a n N a t i o n a l S t a n d a r d f o r I n f o r m a t i o n T e c h n o l o g y)により規格化が進められている。このF C-S B 2には、大きく分けて次の2つの特徴がある。

【0020】1つ目として、従来のメインフレームのプロトコルとは異なり、I/Oの要求処理(1 C C Wチェーン)実行中に、中央処理装置と記憶制御装置間の論理的な接続パス(以下、論理パス)を占有することをせず、同一論理パス上で同時に複数の論理V O Lに対するI/O要求を実行出来るという点である。2つめの特徴としては、中央処理装置は、記憶制御装置とのインターロックを取らずに、コマンド及びデータをパイプライン的に発行出来るという点である。F C-S B 2では、例えばW Rコマンド発行時では、記憶制御装置からW Rコマンドに対するコマンド応答が送信されてこなくても、中央処理装置は、W Rコマンドのデータを記憶制御装置に対して送信する事が可能である。更に、当該コマンドに対するステータスフレームを受信しなくても、中央処理装置は、次コマンド及びデータの発行が可能である。この様に、F C-S B 2では、中央処理装置と記憶制御装置とが非同期にそれぞれコマンドの処理を実行していくプロトコルとなっている。

【0021】以上、説明した様な特徴をもつF C-S B 2プロトコルは、特に、長距離・高負荷接続時にシステムのスループットを低下させないという点で非常に効果的なプロトコルである。

【0022】

【発明が解決しようとする課題】F C-S B 2プロトコルの様な、中央処理装置と記憶制御装置間のインターロック軽減を実現するプロトコルにより接続された中央処理装置及び記憶サブシステムでは、長距離・高負荷接続時におけるスループットの低下を抑えることが出来る。しかし、記憶制御装置は受領したコマンドを受領順に逐次処理していくことには変わりはない。従って、途中のコマンド処理において、処理対象のデータのキャッシュミスが発生した場合、前述の様にキャッシュメモリ上に必要なデータをステージングする間、中央処理装置と記憶制御装置間のデータ転送処理は実行出来ない。この結果、レスポンスタイムの増加を招く。特に、データベースアクセスに代表される様な、ランダムアクセス処理においては、シーケンシャルアクセス時の先読みステージング動作を行なえないため、キャッシュミス契機が多くなる可能性が高くなる。また、あるコマンドでキャッシュミスが発生した場合、そのコマンドからリトライする動作となるため、中央処理装置から受領したコマンド及びデータのうち、キャッシュミスが発生したコマンド以降のコマンド及びデータは一度破棄し、再度中央処理装置から受領し直す必要がある。このコマンド及びデータの再受領は長距離接続時には、かなりのオーバーヘッドとなる。

【0023】本発明の目的は、ランダムアクセス時のキャッシュミス発生時において、レスポンスタイムの増加を防ぎ、F C-S B 2プロトコルの様にパイプライン的に発行されたコマンド及びデータを効率よく処理することが可能な技術を提供することにある。

【0024】本発明の他の目的は、上位装置と記憶サブシステムとが、記憶サブシステム側からの応答とは非同期に、上位装置が複数のコマンド及びデータのチェーンで構成されるI/O要求を記憶サブシステムに対して発行するI/Oインタフェースにて接続された構成において、ランダムアクセス時のキャッシュミス発生時のレスポンスタイム削減によるスループット向上を実現することにある。

【0025】本発明の他の目的は、上位装置と記憶サブシステムとが、F C-S B 2プロトコルにて接続された構成において、ランダムアクセス時のキャッシュミス発生時のレスポンスタイム削減によるスループット向上を実現することにある。

【0026】

【課題を解決するための手段】配下に複数の記憶装置を有し、内部にキャッシュメモリを有する記憶制御装置を備えた記憶サブシステムと、この記憶サブシステムにア

クセスする中央処理装置とを含み、記憶制御装置からの応答とは非同期に、中央処理装置が複数のコマンド及びデータのチェーンで構成されるI/O要求を記憶制御装置に対して発行するI/Oインタフェースプロトコルにより、中央処理装置と記憶制御装置とが接続されている情報処理システムにおいて、本発明の記憶サブシステムは、中央処理装置から受領した複数コマンド及びデータの受領順には依存せず、受領コマンドの処理順序を決定して実行する手段を有するものである。

【0027】また、上記構成の情報処理システムにおいて、本発明の記憶サブシステムは、中央処理装置から受領した複数コマンドのうち、処理対象のデータがキャッシュメモリ上に存在するコマンドについては、中央処理装置とのデータ転送を実行し、処理対象のデータがキャッシュメモリ上に存在しないコマンドの処理対象データを、記憶装置からキャッシュメモリへ読み出す処理と、中央処理装置とのデータ転送処理とを並行して実行する手段を有するものである。

【0028】また、上記構成の情報処理システムにおいて、本発明の記憶サブシステムは、1つの論理ボリュームのデータを複数の記憶装置に分割・配置して格納するRAID構成をとっている場合において、中央処理装置から受領した複数コマンドをまとめてコマンド群とし、コマンド群のうちの複数コマンドの処理対象データが、キャッシュメモリ上に存在しない場合、各コマンドの処理対象データが格納されている各記憶装置に対して、コマンド群の各コマンドの処理対象データのキャッシュメモリへの読み出し処理を複数並行して起動し、この読み出し処理が完了した順にコマンド群の各コマンドに関する中央処理装置とのデータ転送と読み出し処理を並行して実行する手段を有するものである。

【0029】また、上記構成の情報処理システムにおいて、本発明の記憶サブシステムは、中央処理装置から受領した複数コマンドを纏めてコマンド群とし、コマンド群のうちの複数コマンドの処理対象データがキャッシュメモリ上に存在しない場合、各コマンドの処理対象データが格納されている各記憶装置の稼働率に応じて、記憶装置からキャッシュメモリへの処理対象データの読み出し処理を複数並行して起動し、読み出し処理が完了したことから、中央処理装置とのデータ転送を他の読み出し処理と並行して実行する手段を有するものである。

【0030】また、上記構成の情報処理システムにおいて、本発明の記憶サブシステムは、中央処理装置から受領した複数コマンドを纏めて第1のコマンド群とし、第1のコマンド群のうちの複数コマンドの処理対象データが、キャッシュメモリ上に存在しない場合、第1のコマンド群のうち、対象データが同一の記憶装置に存在しているコマンドを記憶装置へアクセスするアドレスの昇順に纏めて第2のコマンド群とし、第2のコマンド群単位で記憶装置からキャッシュメモリへの記憶装置内データ

の読み出し範囲を決定して読み出し処理を起動し、読み出し処理が完了したのちから中央処理装置へのデータ転送を他の読み出し処理と並行して実行する手段を有するものである。

【0031】また、記憶サブシステムと中央処理装置との間を接続しているI/Oインタフェースプロトコル制御方式において、本発明のI/Oインタフェースプロトコル制御技術は、中央処理装置は、当該中央処理装置が発行したコマンド及びデータの発行順には依存しない順で、記憶制御装置からデータ及びコマンド終了報告を受領し、受領したデータ及びコマンド終了報告に対応するコマンドを特定し、データ及びコマンド終了報告を特定したコマンドに対する応答フレームとして処理する手段を有するものである。

【0032】また、このI/Oインタフェース制御方式において、本発明のI/Oインタフェースプロトコル制御技術は、記憶制御装置は、中央処理装置から受領したコマンド及びデータのチェーンの途中でエラー及びリトライ要因が発生した場合、I/O要求を中断せずに継続して、受領した他の実行可能なコマンド及びデータに関して中央処理装置とのデータ転送を実施し、I/O要求を構成する全コマンド処理分の終了状態を纏めて、1つのI/O要求処理完了報告として中央処理装置へ報告する手段を備え、中央処理装置は、I/O要求処理完了報告を受領後、各コマンドの終了状態を認識し、エラーあるいはリトライ要求のあった各コマンドに対してのみリカバリ処理を実行する手段を備えたものである。

【0033】また、このI/Oインタフェース制御方式において、本発明のI/Oインタフェースプロトコル制御技術は、中央処理装置は、当該中央処理装置が発行したコマンド及びデータの発行順には依存しない順で、記憶制御装置からデータ及びコマンド終了報告を受領する手段と、記憶制御装置からI/O要求を構成する全コマンド分に対するコマンド終了報告を受領する迄、I/O要求処理を中断せずに継続して実行する手段と、I/O要求を構成する全コマンド分のコマンド処理終了報告を受領後、受領したコマンド処理終了報告の中でエラーあるいはリトライ要求があったコマンドに対してのみ、リカバリ処理を実行する手段とを有するものである。

【0034】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【0035】まず、本発明を説明するにあたり、本発明を用いた中央処理装置及び記憶サブシステムの構成について、図2を用いて説明する。

【0036】本実施の形態における情報処理システムは、中央処理装置1、ディスク制御装置2、及びディスク装置3から構成されている。ディスク制御装置2は、中央処理装置1とは、たとえばFC-SB2プロトコル(FIBRE CHANNEL Single-Byte

Command Code Sets-2 Mapping Protocol)が実装されたI/Oインタフェース100により接続されている。ディスク装置3には複数の物理ドライブ32が配置されており、データが格納されている。ディスク装置3は、ディスク制御装置2に対してドライブインタフェース200にて接続されている。

【0037】まず、本実施の形態の中央処理装置1の構成例について説明する。中央処理装置1には、ユーザプログラムの要求を実行するアプリケーション部11、アプリケーション部11が発行した入出力要求を受け、実際のデータの入出力命令のためのCCWを作成したり、CCWデータの管理を行なうデータ管理部12、CCWやデータ、制御情報などが格納されている主記憶14、及びディスク制御装置2との間の入出力制御を行うチャネル制御部13から構成されている。

【0038】主記憶14には、CCW情報が格納されているCCW情報格納域142、CCW情報格納域142の先頭アドレスが格納されているCCW先頭アドレス格納域141、CCWのデータが格納されているData格納域143、そしてI/O要求を管理するためのI/O管理領域144がある。これらの詳細については、後述する。

【0039】次に、本実施の形態のディスク制御装置2の構成例について説明する。送受信バッファメモリ21は、中央処理装置1との間で送受信されるコマンド及びデータを一時的に格納するメモリである。対チャネルプロトコル制御部22は、前述のFC-SB2プロトコルを制御する。コマンド処理部23は、中央処理装置1から受領したコマンドをデコードし、コマンド処理を行う。キャッシュメモリ24は、送受信バッファメモリ21と、ディスク装置3内の各物理ドライブ32との間のデータ転送において、データを一時的に格納するためのメモリである。制御メモリ25には、コマンド処理用の制御情報、物理ドライブ32とキャッシュメモリ24との転送を制御するための情報及び、後述の本実施の形態における各種制御を実施する上で必要な各種情報が格納されている。各テーブルの内容、用途については後述する。ディスクドライブ制御部26は、ディスク制御装置2に接続されているディスク装置3とのインタフェース制御を行う。

【0040】ディスク装置3には、ディスク制御装置2とのインタフェース制御を行う、ディスク装置インタフェース制御部31及び、データを格納する物理ドライブ32が複数配置されている。本実施の形態では、RAID5構成を用いてデータを格納しているが、これについて、図3を用いて説明する。

【0041】中央処理装置1で扱うデータボリュームを論理ボリュームと言うが、図3にそのうちの1論理ボリューム4を示す。シリンダ(以下、CYL)及びヘッド

(以下、HD)の組み合わせで決まるエリアをトラックという。論理ボリューム4は、複数のトラックから構成されている。本実施の形態では、論理ボリュームの各トラックを複数の物理ドライブ32に分割して配置する。図3の場合、論理ボリューム4内のトラック41はドライブ#0の物理ドライブ32に、トラック42はドライブ#1の物理ドライブ33に、そしてトラック43は、ドライブ#2の物理ドライブ34に格納する。そして、トラック41~43からパリティデータを生成し、パリティデータを示すトラック44をドライブ#3の物理ドライブ35に格納する。この時、パリティデータの生成単位であるトラック41~44の横一列のトラックの並びをストライプ列という。本実施の形態は、RAID5構成であるので、途中でパリティデータを格納する物理ドライブの位置がサイクリック的に移動していく。本実施の形態では、8ストライプ列毎にパリティデータの格納物理ドライブを代えていく。また、物理ドライブ32は、固定長ブロック(以下、LBA)に分割されており、CCWのデータはこの固定長に分割して格納される。

【0042】次にI/O要求処理の基本的な流れについて、図2を用いて説明する。

【0043】まず、中央処理装置1にあるアプリケーション部11において、ディスク装置3に格納されている論理ボリューム(以下、論理VOL)データに対する入出力要求が作成され、この要求がデータ管理部12に発行される。データ管理部12は、発行された入出力要求を実行するために、CCWチェーンを作成する。作成されたCCWは、主記憶14内のCCW情報格納域142に格納され、CCWのデータは、Data格納域143に格納される。また、CCW先頭アドレス格納域141に、CCW情報格納域142の先頭アドレスを格納しておく。データ管理部12はCCWを作成したら、チャネル制御部13に対し、CCWの実行を要求する。CCW実行要求を受けたチャネル制御部13は、CCW先頭アドレス格納域141を参照し、CCW情報格納域142の先頭アドレスを取得して、CCWを得る。そして、CCW情報格納域142に格納されているCCWのコマンドをディスク制御装置2に逐次発行していく。また、WR系コマンドの場合には、発行したコマンドに関連するデータは、Data格納域143に格納されているので、ディスク制御装置2に送信する。また、この時、データ管理部12はI/O管理領域144に図4に示すI/O要求共通情報1440以下の内容を登録する。その後、逐次CCWをディスク制御装置2に発行する度に必要なCCW情報をI/O管理領域144のCCW管理情報1441に登録する。

【0044】一方、チャネル制御部13から受領したCCWのコマンド及びデータは、ディスク制御装置2の送受信バッファメモリ21に格納される。対チャネルプロ

トコル制御部22は、I/O管理テーブル256に図8に示すI/O要求共通情報2560以下の内容を登録し、管理する。また、パイプライン的に受領した後続のCCWについては、それぞれのCCWに対応したCCW管理情報2561に登録し、管理する。対チャンネルプロトコル制御部22は、各CCWコマンド受領毎に、コマンドの受領をコマンド処理部23に通知する。コマンド処理部23は、受領したコマンドをデコードし、キャッシュデータ管理テーブル252を参照して、処理対象データがキャッシュメモリ24上に存在するか否かをチェックする。キャッシュデータ管理テーブル252には、対象の論理VOL#/CYL#/HD#/レコード#のデータがキャッシュメモリ24上に存在しているか否かの情報や、存在している場合のキャッシュメモリ24上のアドレス、データの属性などの情報が格納されている。キャッシュメモリ24上にデータが存在（以下、キャッシュヒット）した場合、コマンド処理部23は、キャッシュメモリ24と送受信バッファメモリ21間のデータ転送を実施する。一方、対チャンネルプロトコル制御部22は、送受信バッファメモリ21と、チャンネル制御部13との間のデータ転送を実施する。

【0045】一方、キャッシュメモリ24上に処理対象データが存在しなかった（以下、キャッシュミス）場合、対象データをそのデータが格納されている物理ドライブ32からキャッシュメモリ24上に読み出す（以下、ステージング）処理が必要となる。このステージング処理は、物理ドライブ毎に独立して動作可能である。処理対象データが格納されているディスク装置3内の物理ドライブ番号は、論理-物理アドレス変換テーブル251を参照する事で取得する。このテーブルには、論理VOL#/CYL#/HD#/レコード#から決定される処理対象のレコードが、どの物理ドライブのどのLBAに格納されているか、が示されている。論理-物理アドレス変換テーブル251よりステージング対象の物理ドライブ#を取得したら、図5に例示されるドライブ別ステージング要求キューテーブル253にステージング要求内容を登録する。ドライブ別ステージング要求キューテーブル253はFIFO構造になっており、ステージングの要求順に登録される。登録する内容については、図5に例示されている通りである。ここで、ステージング要求を登録する際、各ドライブ毎にユニークな要求IDを付与する。ドライブ別ステージング要求キューテーブル253にステージング要求を登録後、ステージング起動キューテーブル254に図6に例示する内容を登録する。また、更に図8のI/O管理テーブル256の当該CCW#に対応するCCW管理情報のCCW管理情報2561に“ステージング完了待ち”を設定する。

【0046】一方、ディスクドライブ制御部26は、図6のステージング起動キューテーブル254を定期的に参照し、ステージング要求が登録されたら、ステージン

グ対象の物理ドライブ#からそのドライブ別ステージング要求キューテーブル253を参照して、ステージング内容を取得し、ステージング要求をディスク装置インタフェース制御部31に対して発行する。要求を受けたディスク装置インタフェース制御部31は、目的の物理ドライブから指示されたステージング開始LBA#（以下、SLBA#）から必要分のデータをキャッシュメモリ24に転送する。ステージング処理が終了すると、ディスク装置インタフェース制御部31からディスクドライブ制御部26に対し、ステージング終了の報告を行なう。報告を受けたディスクドライブ制御部26は、ステージング完了報告キューイングテーブル255にステージング処理が終了したドライブ#/要求IDをキューイングする。

【0047】コマンド処理部23はステージング完了報告キューイングテーブル255を参照し、ステージング終了を検出すると、ステージング完了報告キューイングテーブル255のドライブ#及び要求IDを基に、ドライブ別ステージング要求キューテーブル253を参照して、ステージングが完了したI/O要求#/CCW#を取得する。コマンド処理部23は、取得したI/O要求#/CCW#を基にI/O管理テーブル256を参照し、当該CCW#に対応するCCW管理情報のCCW管理情報2561の“ステージング完了待ち”を“CCW処理中”に変化させ、コマンド処理を再開する。後はキャッシュヒット時の動作と同じになる。

【0048】以下、本発明の実施の形態についてさらに詳細に説明する。

【0049】本発明を実施したI/Oシーケンスの例を図1に示す。図1では、中央処理装置1からディスク制御装置2に対し、CMD101~CMD103が発行され、コマンド処理部23において、CMD101のCCW1及びCMD102のCCW2がキャッシュミスだったと認識した（キャッシュミス、ヒットの判定については前述の通り）。この時、前述の方式により、CMD101のCCW1の処理対象データのステージング処理を起動し、続いてCMD102のCCW2の処理データのステージング処理も並行して起動する（2010、2011）。続いて、コマンド処理部23は、後続のCMD103（CCW3）がキャッシュヒットだったため、コマンド処理部23は、CCW3の処理を行い、CCW3のデータであるDATA104を中央処理装置1に送信する。CCW3のデータを送信した後、図1では、CCW1のステージング要求に対するステージング完了通知2012が報告されているため、コマンド処理部23はCCW1の処理を実行し、中央処理装置1に対してCCW1のDATA105を送信する。その後、CCW2のステージング要求に対するステージング完了通知2013が報告されてきたため、コマンド処理部23はCCW2の処理を実行し、中央処理装置1に対してCCW2の

DATA106を送信する。最後にSTATUS107を送信する。

【0050】ここで、CCW3の処理が終了した時点で、まだCCW1に対するステージング完了通知2012が報告されていなかった場合は、ディスク制御装置2は中央処理装置1との論理的な接続を切断し、ステージング完了通知2012があつてから、再度中央処理装置1との論理的な接続を回復し、CCW1の処理を実行してもよい。また、中央処理装置1との論理的な接続は切り離さず、ステージング完了通知2012の報告を待ってからCCW1の処理を実行してもよい。図1では、CCW1に対するステージング完了通知(2012)の方が先に発生したが、CCW2に対するステージング完了通知(2013)の方が先に発生した場合は、先にCCW2の処理を実行してもよい。

【0051】もし、中央処理装置1から受領した複数CCWコマンドがキャッシュミスだった場合は、各CCW毎にドライブ別ステージング要求キューテーブル253及び、ステージング起動キューテーブル254にステージング要求がキューイングされる。ステージング処理は物理ドライブ毎のため、ステージングの起動順にステージングが完了するとは限らない。この場合、ステージングが完了した順にステージング完了報告キューイングテーブル255にキューイングされ、このキューイング順でステージング完了報告が行われる。コマンド処理部23はステージング完了報告キューイングテーブル255を参照してコマンド処理を実行するため、結果として、ステージングが完了した順にコマンド処理を実行することになる。

【0052】上記の処理は、次の様にしてもよい。しきい値テーブル259(図11)に受領CCW数しきい値2592を持ち、中央処理装置1から受領したCCW数が、この受領CCW数しきい値2592に達した場合、そこまでのCCW迄をCCW群として1つの処理対象範囲とし、図12に示すフローにて制御してもよい。

【0053】すなわち、図12のフローチャートにおいては、まず、受領CCW数を0に初期化した後(ステップ120001)、中央処理装置1からのCCW到来を待ち(ステップ120002、ステップ120003)、CCWを受領したら、受領CCW数を加算し(ステップ120004)、当該CCWの対象データのヒット/ミス判定を実行し(ステップ120005)、ヒットの場合は当該CCWの処理を実行し(ステップ120006)、ミスの場合は、当該対象データのドライブからキャッシュメモリ24へのステージング処理を実行する(ステップ120007)。

【0054】さらに、後続のCCWの有無を判別し(ステップ120008)、後続有りの場合には、受領済みのCCW数は、受領CCWしきい値以下か否かを判定し(ステップ120009)、しきい値以下の場合は、ス

テップ120002以降のCCWの受け付け処理を反復する。

【0055】ステップ120008で後続なしの場合、あるいはステップ120009でしきい値以上の場合には、受領した全CCWがキャッシュミスか否かを判定し(ステップ120010)、キャッシュミスの場合には、ステージング処理待ち状態に移行する(ステップ120011)。

【0056】ステップ120010で受領した全CCWがキャッシュミスではない場合には、STATUS Frame送信処理を実行する。

【0057】上述のステージング処理待ち状態(ステップ120011)では、キャッシュミスのCCWの対象データのステージング処理の終了を監視し(ステップ120012、ステップ120013)、ステージング処理の終了したCCWがある場合には、当該ステージング処理の終了したCCWの処理を実行し(ステップ120014)、キャッシュミスの全CCWの対象データの処理が完了したか否かを判別し(ステップ120015)、未完の場合には、ステップ120012以降を反復し、完了の場合には、STATUS Frame送信処理を実行する。

【0058】中央処理装置1から受領した複数CCWにおいて、キャッシュミスが発生し、ステージング処理の起動が必要な場合、図13に示すフローの様に制御してもよい。すなわち、まず、受領CCW数を0に初期化した後(ステップ130001)、中央処理装置1からのCCW到来を待ち(ステップ130002、ステップ130003)、CCWを受領したら、受領CCW数を加算し(ステップ130004)、当該CCWの対象データのヒット/ミス判定を実行し(ステップ130005)、ヒットの場合は当該CCWの処理を実行し(ステップ130006)、ミスの場合は、当該対象データのドライブからキャッシュメモリ24へのステージング処理を実行する(ステップ130007)。

【0059】さらに、後続のCCWの有無を判別し(ステップ130008)、後続有りの場合には、受領済みのCCW数は、受領CCWしきい値以下か否かを判定し(ステップ130009)、しきい値以下の場合は、ステップ130002以降のCCWの受け付け処理を反復する。

【0060】ステップ130008で後続なしの場合、あるいはステップ130009でしきい値以上の場合には、キャッシュミスのCCW無しか判定し(ステップ130010)、無しの場合には、STATUS Frame送信処理を実行する。

【0061】有りの場合には、キャッシュミスのCCWの各データのキャッシュメモリへのステージング処理を起動すべき対象ドライブ番号を取得し(ステップ130011)、ステージング対象のCCWは残り一つか判定

し（ステップ130012）、残り一つの場合には当該CCWに関してステージング処理を起動する（ステップ130013）。

【0062】残り一つではない場合には、ドライブ別稼働管理テーブル257を参照して（ステップ130011）決定した対象ドライブのうち、一番稼働率の高いドライブ番号を選択し（ステップ130014）、選択した以外のドライブに関してステージング処理を起動し（ステップ130015）、ステージング起動未完かを調べ（ステップ130016）、未完の場合には、ステップ130012以降を反復し、完了の場合には、ステージング完了待ち処理へ移行する。

【0063】なお、図13において、ステップ130014で、ステージング対象ドライブのうち、一番稼働率が高いドライブ番号を取得したが、これは図9のドライブ別稼働管理テーブル257のカレントエリア#2570（エリア#0、エリア#1のいずれか）を参照し、カレントなエリア（統計的なデータを採集中のエリア）の各ステージングが必要なドライブ#のアクセス回数をそれぞれ比較することで得られる。ディスク装置3のディスク装置インタフェース制御部31は、各物理ドライブにアクセスする度に、カレントエリア#2570が示しているエリアのアクセス回数を1インクリメントする。なお、ドライブ別稼働管理テーブル257のカレントエリア#2570は、一定時間超過すると、エリア#0とエリア#1の間で交互に逆エリアを指し示す様になっている。

【0064】また、図15及び図16に示すフローチャートによって、複数のCCWに対するステージングを1つのステージング要求として纏めて起動してもよい。

【0065】すなわち、図15の処理では、まず上述の図13のステップ130001～ステップ130009までと同様の処理の後、受領CCWしきい値分のCCWのヒットミス判定およびキャッシュヒットしたCCWの処理を実行し（ステップ150001）、全CCWキャッシュミスでないか判定し（ステップ150002）、全CCWキャッシュミスでない場合は、STATUS Frame送信処理を実行する。

【0066】一つでもキャッシュミスのCCWがある場合、ドライブ別ステージング要求ソートテーブル258に、アクセス対象アドレスが昇順になるようにステージング要求のCCWをキューイングし（ステップ150003）、ドライブ別ステージング要求のまとめステージング処理を実行し（ステップ150004）、ステージング完了待ち処理を移行する（ステップ150005）。

【0067】上述のステップ150004のまとめステージング処理では、ループ変数Iを1に初期化した後（ステップ160001）、変数Nにドライブ別ステージング要求ソートテーブル258にエントリされている

CWW数をセットし、I番目のステージング開始論理ブロックアドレスSLBA(I)を変数SLBAにセットし、I番目のステージング終端論理ブロックアドレスELBA(I)を変数ELBAにセットし（ステップ160002）、I+1がエントリされているCWW数をこえない間（ステップ160003）、ELBA(I)と次のSLBA(I)の間隔（差分）が、まとめステージングLBA数しきい値以下か判別し（ステップ160004）、当該しきい値以下の場合には、ELBA(I)を変数ELBAにセットし（ステップ160005）、ループ変数Iをインクリメントし（ステップ160006）、ステップ160003以降を反復することで、とびとびのステージング範囲を一つの領域にまとめる処理を行う。

【0068】ステップ160003でIがI+1がエントリされているCWW数を超えた場合、またはステップ160004で隣り合うステージング範囲の間隔がまとめステージングLBA数しきい値以上離れている場合には、CCW#1～CCW#IまでのCCW群（一つのCCWの場合もある）に関してステージング要求IDを付与する（ステップ160007）。

【0069】ここで、図15におけるステップ150003では、ステージング対象のドライブ単位にドライブヘアクセスするアドレスの昇順にCCW毎のステージング要求を並べているが、ドライブ別ステージング要求ソートテーブル258にステージング要求をキューイングする際に必ず昇順になるように新規ステージング要求をキューイングする。図16のフローチャートでは、複数のステージング要求をしきい値テーブル259の、纏めステージングLBA数しきい値2591を用いて、1つのステージング要求に置き換える処理を行なっている。これについて、図14を使って詳細に説明する。個々のCCWのステージング範囲が図14に示す状態だったとする。ここで、CCW#nのステージング範囲140005とCCW#(n+1)のステージング範囲140006の隙間（SLBA#(n+1)140003-ELBA#(n)140002）と、纏めステージングLBA数しきい値2591とを比較し、ステージング範囲の隙間の方が小さければ、CCW#nとCCW#(n+1)の両方のステージング範囲を一緒にして、1つのステージング要求として起動する。マージされたステージング範囲のステージング開始位置はSLBA#(n)140001となり、終端はELBA#(n+1)140004となる。しかし、隙間より、纏めステージングLBA数しきい値2591の方が小さければ、ステージング範囲はマージしない。

【0070】次に本実施の形態のI/Oインタフェースの制御方法について説明する。

【0071】本実施の形態では、中央処理装置1から受領したCCWの順にディスク制御装置2はCCWを処理

しない。従って、送信するデータの順も中央処理装置1が発行したCCWの順とは異なる場合がある。本実施の形態の中央処理装置1では、データを受信したチャンネル制御部13は、データフレームに記載のI/O要求#及びCCW#から、I/O管理領域144を参照し、各I/Oの各CCWのデータ格納アドレスを取得し、そのアドレスに受領したCCWのデータを格納する。こうする事で、チャンネル制御部13が発生したCCW順でないデータフレームやステータスフレームを受領しても対応可能である。

【0072】また、ディスク制御装置2において、CCWチェーン途中のCCWでエラーが発生した場合、次の様にしてもよい。本実施の形態のI/Oインタフェースの制御方法では、図17に示す様にステータスフレームの内容を変更し、1つのステータスフレームの中に複数CCW分のステータス情報を含んだものとする。具体的に説明すると、図17のステータスフレームに制御情報170001を新規に設け、この中に、複数CCWの情報を含んでいる多重ステータス報告機能を有しているかを判定するためのビットを持つ。多重ステータス報告時、1つのステータスフレームには複数のCCWに関するステータス情報が含まれており、それぞれにこの制御情報170001が付与される。この時、幾つのCCWのステータス情報が含まれているかを示すために、各CCWの制御情報内に“後続STATUS報告チェーンビット”を設け、このチェーンビットがONの時、後続ステータス報告がありと判断する。チェーンビットがOFFのステータス報告で最後と判断する。

【0073】このステータスフレームを用いた本実施の形態のI/Oプロトコルシーケンスを図18に示す。図18において、CCW2でエラー要因が発生したが、CCW2迄でCCWチェーンを切らず、ディスク制御装置2ではCCW3の実行も行なう。そして、全CCWの実行後、図17に記載のステータスフレームを使って、CCW1~3のステータス情報を一緒に載せて報告する。

【0074】また、このステータスフレームを受領したチャンネル制御部13は、当該I/Oの各CCWの終了状態をI/O管理領域144に記憶する。その後、チャンネル制御部13は当該I/Oの全CCWの終了状態をデータ管理部12に報告する。データ管理部12はI/O管理領域144を参照し、エラー及びリトライ要因が発生したCCWを特定して、必要に応じてリカバリ処理を実行する。

【0075】また、次のようにしてもよい。ディスク制御装置2は、図19に示す様にディスク制御装置2が決定した順でCCWのデータの送信を実施し、各CCWの処理終了毎にステータスフレームを報告する。この時にステータスフレームは図17に示したものでなくてもよい。この場合、各CCW毎に送信されるステータスフレームを受領したチャンネル制御部13側での対応が必要で

ある。チャンネル制御部13は、受領したステータスフレームからI/O要求#/CCW#を取得し、I/O管理領域144に当該CCWの終了状態を格納する。この時、例えば、エラー及びリカバリを示すステータスフレームを受領しても、チャンネル制御部13は、当該I/Oの全CCW分のステータスフレームを受領していない場合は、データ管理部12へCCWチェーン終了の報告は行なわない。当該I/Oの全CCWに対するステータスフレームを受領した後、データ管理部12に対して当該CCWチェーンの終了報告割り込みを行う。後は、前述の通り、データ管理部12が必要に応じてエラー及びリトライ要因が発生したCCWに対するリカバリ処理を実行する。

【0076】以上説明したように、本実施の形態の記憶サブシステム及びI/Oインタフェース制御技術を用いた情報処理システムによれば、FC-SB2のような接続インタフェースにて接続された中央処理装置1と記憶サブシステム（ディスク制御装置2）において、ディスク制御装置2（記憶制御装置）は、対象データがキャッシュヒットして即時実行可能なCCWに関して中央処理装置1と記憶サブシステム間でデータ転送を実行すると同時に、キャッシュミスのCCWに関するデータのステージング処理を並行して実行することが出来るため、たとえば、ランダムアクセス時において、発生し易くなるキャッシュミスによるレスポンスタイムの増加を抑える事が出来るという効果が得られる。

【0077】また、中央処理装置1では、一群のCCWのディスク制御装置2に対する発行順序に関係なく、ディスク制御装置2からのデータ及びCCW終了報告を受領し、受領したデータ及びCCW終了報告に対応する発行済みCCWを特定し、データ及びCCW終了報告を特定したCCWに対する応答フレームとして処理するので、たとえば、複数のCCWの一部がキャッシュミスによるステージング処理のために、他のキャッシュヒットのCCWが実行を待たされる等の不具合が解消され、たとえば、ランダムアクセス時において、発生し易くなるキャッシュミスによるレスポンスタイムの増加を抑える事が出来るという効果が得られる。

【0078】この結果、中央処理装置1と、ディスク制御装置2およびディスク装置3からなる記憶サブシステムとの間におけるI/Oのスループットが向上する。

【0079】以上、本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0080】たとえば、中央処理装置と記憶制御装置とを接続するI/Oシーケンスプロトコルとしては、上述の実施の形態に例示したFC-SB2等に限らず、記憶制御装置からの応答とは非同期に中央処理装置が記憶制御装置に対して複数のコマンドを発行するI/Oインタ

フェースに広く適用することができる。

【0081】

【発明の効果】本発明によれば、ランダムアクセス時のキャッシュミス発生時において、レスポンスタイムの増加を防ぎ、FC-SB2プロトコルの様にパイプライン的に発行されたコマンド及びデータを効率よく処理することができる、という効果が得られる。

【0082】本発明によれば、上位装置と記憶サブシステムとが、記憶サブシステム側からの応答とは非同期に、上位装置が複数のコマンド及びデータのチェンで構成されるI/O要求を記憶サブシステムに対して発行するI/Oインタフェースにて接続された構成において、ランダムアクセス時のキャッシュミス発生時のレスポンスタイム削減によるスループット向上を実現することができる、という効果が得られる。

【0083】本発明によれば、上位装置と記憶サブシステムとが、FC-SB2プロトコルにて接続された構成において、ランダムアクセス時のキャッシュミス発生時のレスポンスタイム削減によるスループット向上を実現することができる、という効果が得られる。

【図面の簡単な説明】

【図1】本発明の一実施の形態である記憶サブシステムの作用の一例を示すシーケンスフローチャートである。

【図2】本発明の一実施の形態である記憶サブシステムを含むデータ記憶システムの構成の一例を示すブロック図である。

【図3】本発明の一実施の形態である記憶サブシステムにおけるディスク装置でのデータ格納方法の一例を示す説明図である。

【図4】本発明の一実施の形態である記憶サブシステムに接続される中央処理装置にて用いられている制御情報の一例を示す説明図である。

【図5】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図6】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図7】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図8】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図9】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図10】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図11】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置にて用いられている制御情報の一例を示す説明図である。

【図12】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置の作用の一例を示すフローチャートである。

【図13】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置の作用の一例を示すフローチャートである。

【図14】本発明の一実施の形態である記憶サブシステムの作用の一例を示す説明図である。

【図15】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置の作用の一例を示すフローチャートである。

【図16】本発明の一実施の形態である記憶サブシステムを構成する記憶制御装置の作用の一例を示すフローチャートである。

【図17】本発明の一実施の形態であるI/Oインタフェースの制御方法におけるI/Oプロトコルの一例を示す説明図である。

【図18】本発明の一実施の形態である情報処理システムの作用の一例を示すシーケンスフローチャートである。

【図19】本発明の一実施の形態である情報処理システムの作用の一例を示すシーケンスフローチャートである。

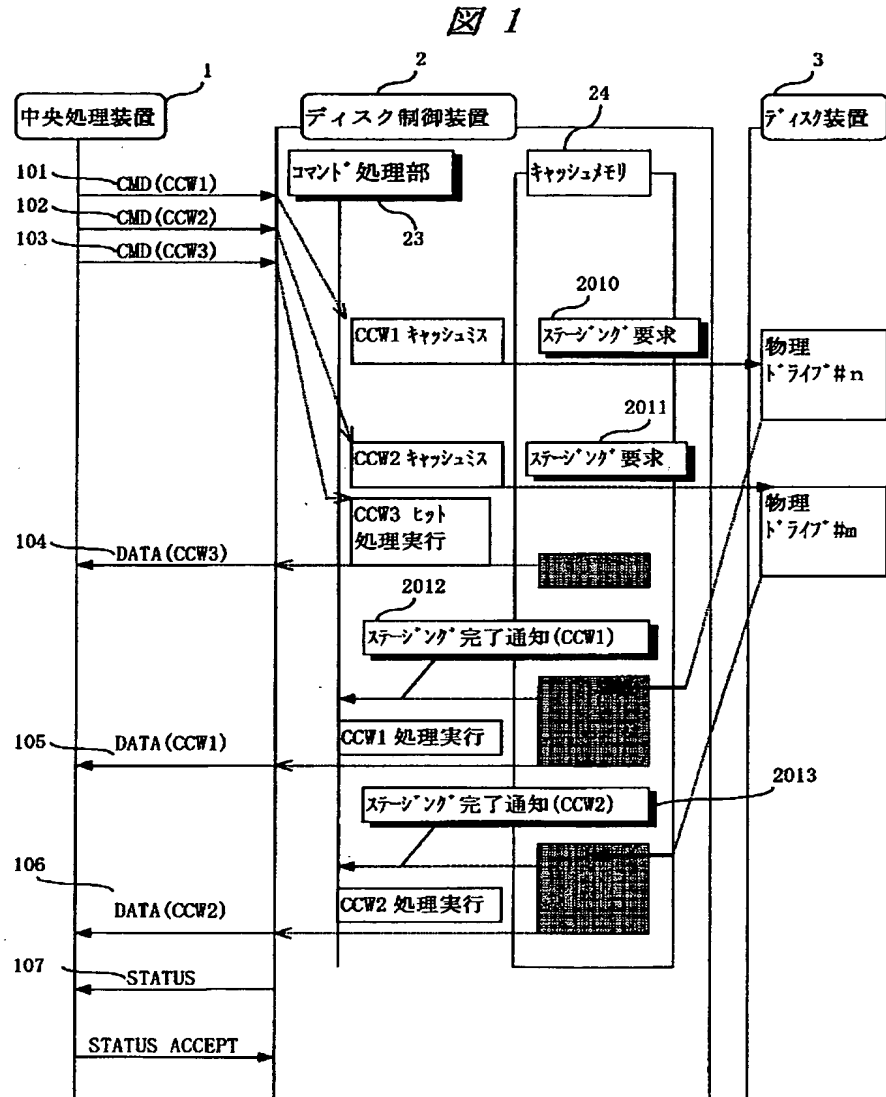
【符号の説明】

1…中央処理装置、101～107…フレーム、11…アプリケーション部、12…データ管理部、13…チャネル制御部、14…主記憶、141…CCW先頭アドレス格納域、142…CCW情報格納域、143…Data格納域、144…I/O管理領域、1440…I/O要求共通情報、1441…CCW管理情報、2…ディスク制御装置（記憶制御装置）、21…送受信バッファメモリ、22…対チャネルプロトコル制御部、23…コマンド処理部、24…キャッシュメモリ、25…制御メモリ、2010…ステージング要求、2011…ステージング要求、2012…ステージング完了通知（CCW1）、2013…ステージング完了通知（CCW2）、251…論理-物理アドレス変換テーブル、252…キャッシュデータ管理テーブル、253…ドライブ別ステージング要求キューテーブル、254…ステージング起動キューテーブル、255…ステージング完了報告キューイングテーブル、256…I/O管理テーブル、2560…I/O要求共通情報、2561…CCW管理情報、257…ドライブ別稼働管理テーブル、2570…カレントエリア#、258…ドライブ別ステージング要求ソートテーブル、259…しきい値テーブル、2590…ステージング対象コマンド数しきい値、2591…纏めステージングLBA数しきい値、2592…受領C

CW数しきい値、26…ディスクドライブ制御部、3…ディスク装置（記憶装置）、31…ディスク装置インタフェース制御部、32…物理ドライブ、4…論理ボリューム、41～44…トラック、120001～120015…処理ステップ、130001～130016…処理ステップ、140001…CCW#(n)ステージング範囲先頭、140002…CCW#(n)ステージング範囲終端、140003…CCW#(n+1)ステージング範囲先頭、140004…CCW#(n+1)ス

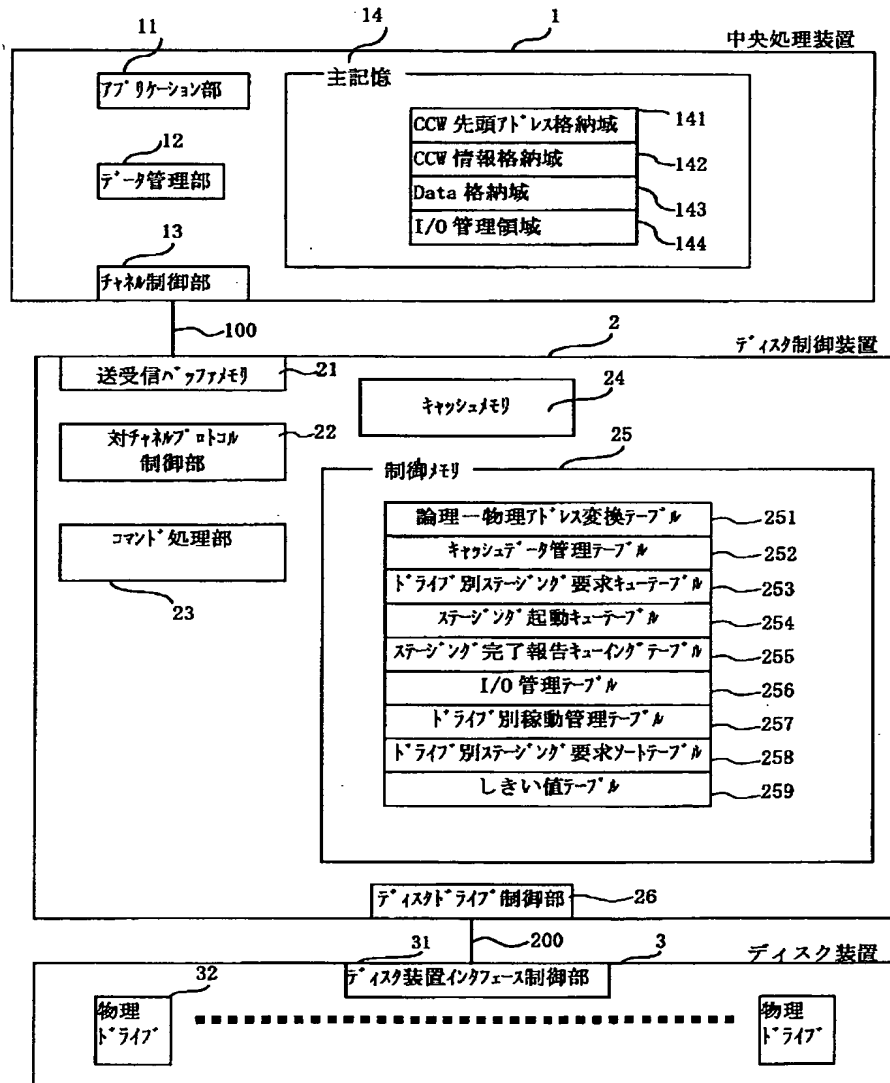
テージング範囲終端、140005…CCW#nステージング範囲、140006…CCW#(n+1)ステージング範囲、150001～150005…処理ステップ、160001～160007…処理ステップ、170001…ステータスフレーム内制御情報、180001～180008…フレーム、190001～190010…フレーム、100…I/Oインタフェース、200…ドライブインタフェース。

【図1】



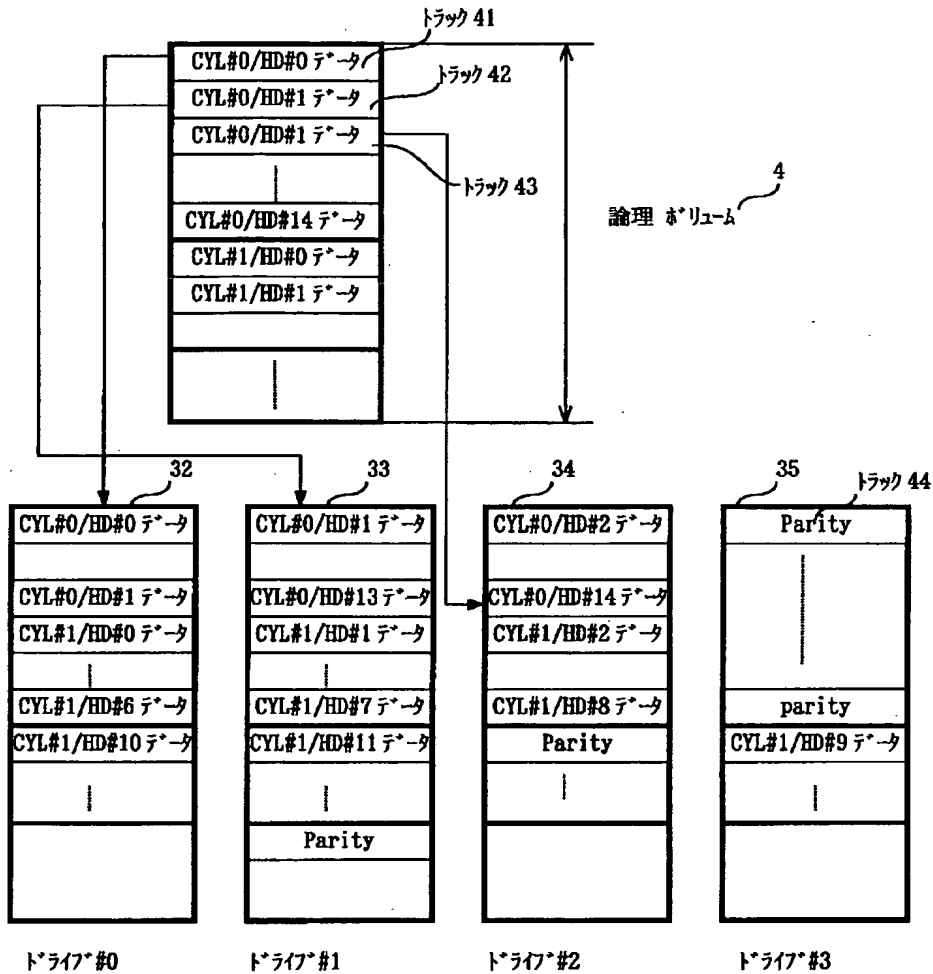
【図2】

図 2



【図3】

図 3



【図6】

図 6

254
スキャンク起動キューデータ

0	1	2	3 (Byte)
Reserved	CNT	IP	OP
要求 1			
要求 2			
要求 3			
要求 n			

CNT: エントリされている要求の数。
 IP: 一番最近にエントリされた要求のエリアを示す。
 OP: 一番古くエントリされた要求のエリアを示す。
 要求の内容: ドライブ#/要求 ID#

【図7】

図 7

255
スキャンク完了報告キューデータ

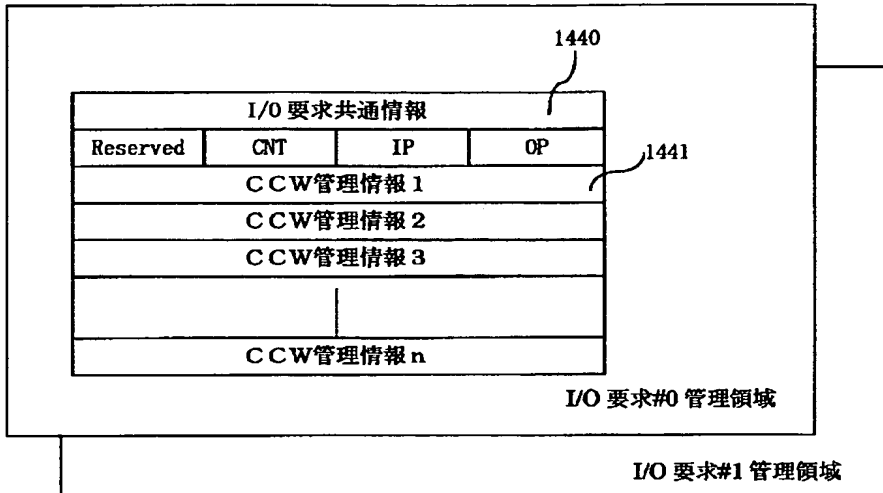
0	1	2	3 (Byte)
Reserved	CNT	IP	OP
完了報告 1			
完了報告 2			
完了報告 3			
完了報告 n			

CNT: エントリされている完了報告の数。
 IP: 一番最近にエントリされた完了報告のエリアを示す。
 OP: 一番古くエントリされた完了報告のエリアを示す。
 完了報告の内容: ドライブ#/要求 ID#/終了状態

【図4】

図 4

144



CNT: エントリされている CCW 管理情報の数。
 IP: 一番最近にエントリされた CCW 管理情報のエリアを示す。
 OP: 一番古くエントリされた CCW 管理情報のエリアを示す。

I/O 要求共通情報: I/O 処理状態情報

I/O 状態情報: 全 CCW の STATUS 受領状態
 (全 STATUS 受領済み or 未受領)

CCW 管理情報: CCW#/コマンドコード/CCW 状態情報/ステージング要求 ID#
 /Data 格納先主記憶上アドレス

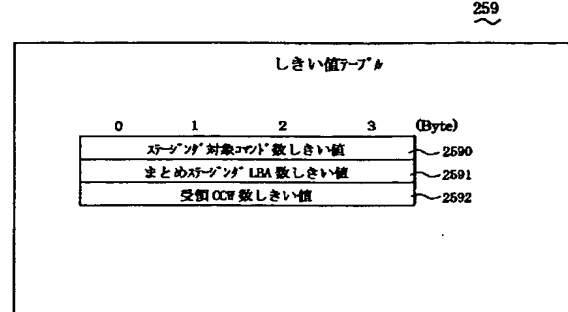
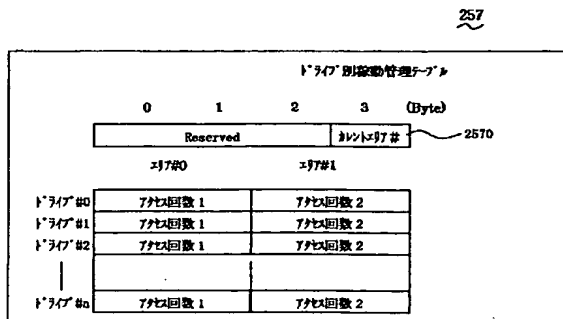
CCW 状態情報: CCW 処理中/STATUS 受領待ち

【図9】

【図11】

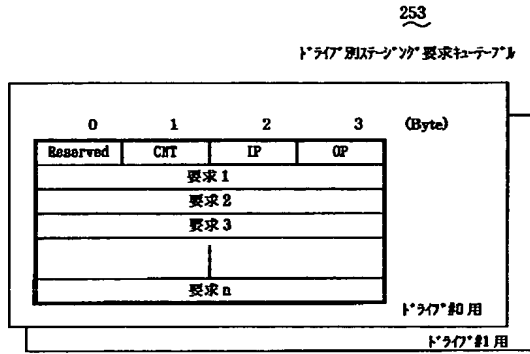
図 9

図 11



【図5】

図 5



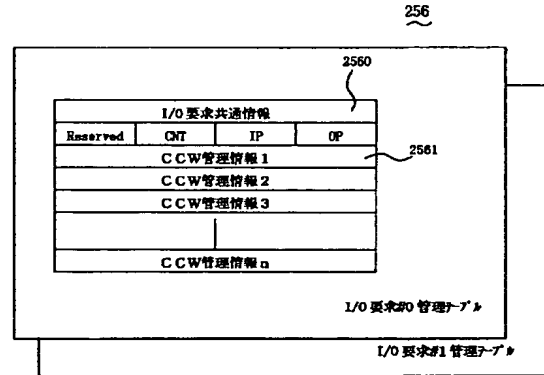
CNT: エントリされている要求の数。
IP: 一番最近にエントリされた要求のエリアを示す。
OP: 一番古くエントリされた要求のエリアを示す。

要求内容:

要求 ID/(I/O 要求#)/CCW#/転送開始 LBA#/転送終了 LBA#/転送 Byte 数
/REX #/CZ

【図8】

図 8



CNT: エントリされている CCW 管理情報の数。

IP: 一番最近にエントリされた CCW 管理情報のエリアを示す。

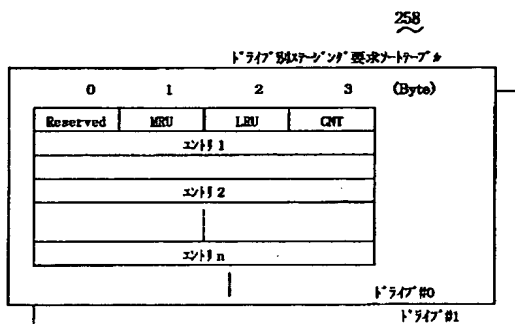
OP: 一番古くエントリされた CCW 管理情報のエリアを示す。

I/O 要求共通情報: I/O 要求 #

CCW 管理情報: CCW#/コマンド/CCW 状態情報/ステージング要求 ID#/
CCW 状態情報: CCW 処理中/ステージング完了待ち

【図10】

図 10



MRU: 一番最近のエントリのエリアを示す。

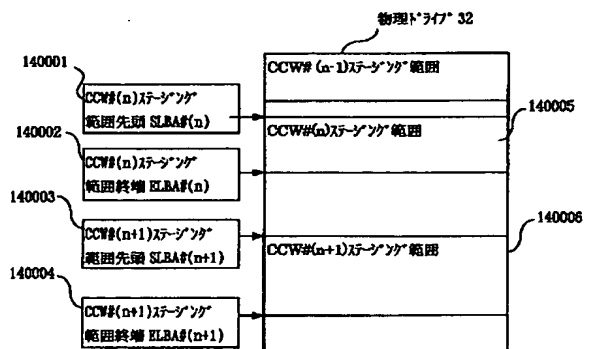
LRU: 一番古いエントリのエリアを示す。

CNT: エントリ数。

エントリの内容: 要求 ID#/(I/O 要求#)/CCW#/転送開始 LBA#/転送終了 LBA#
/転送 Byte 数

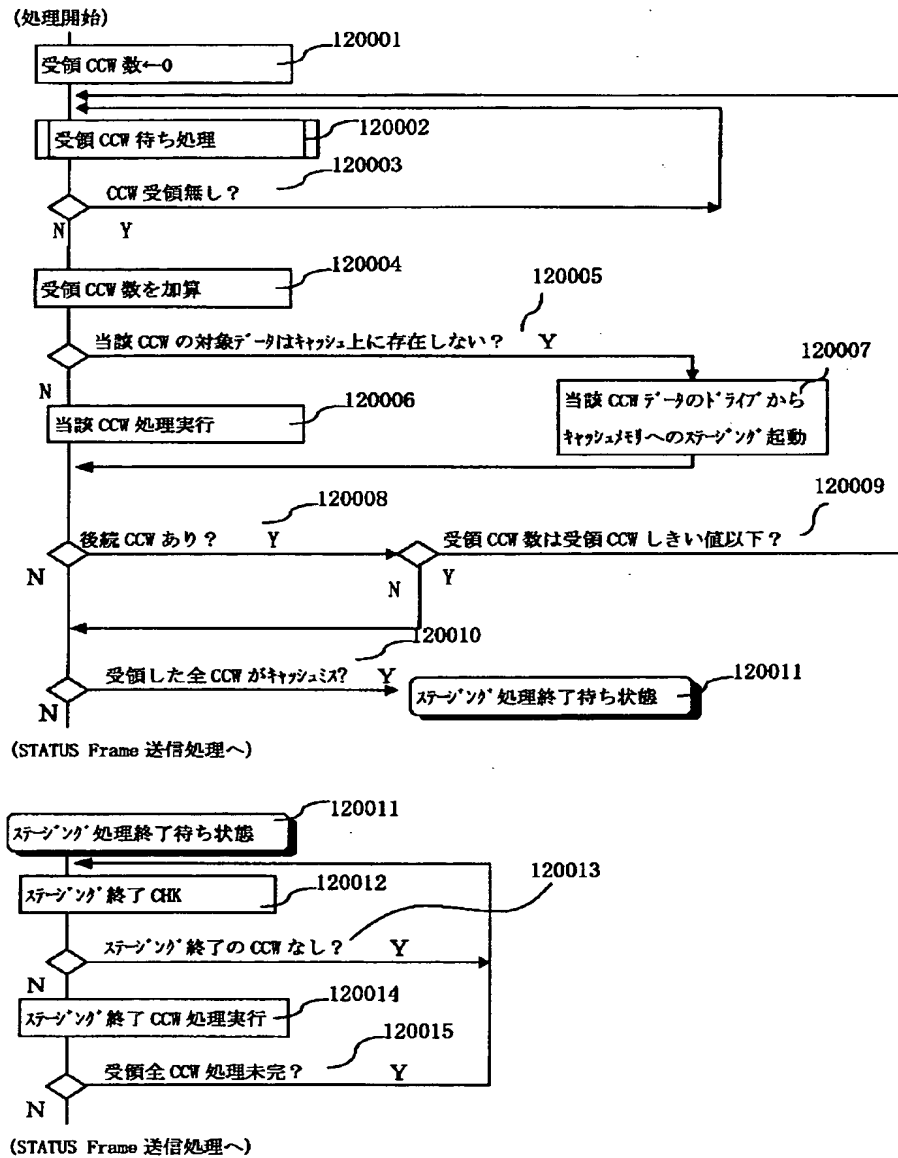
【図14】

図 14

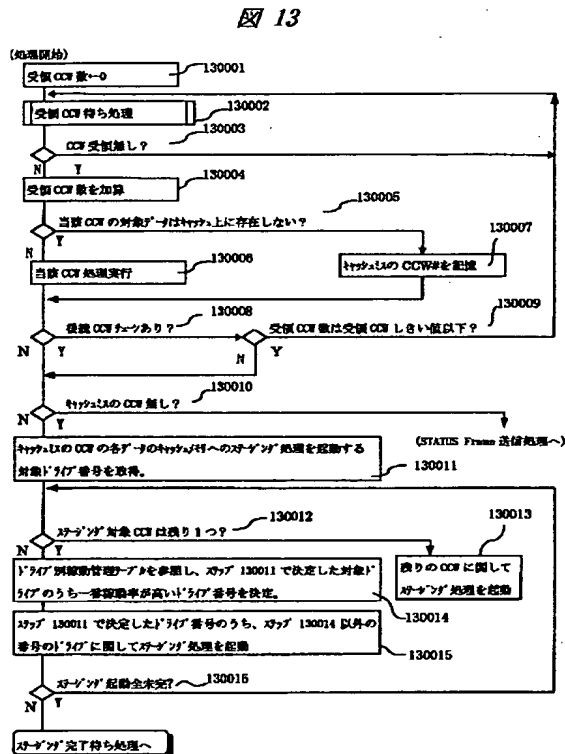


【図12】

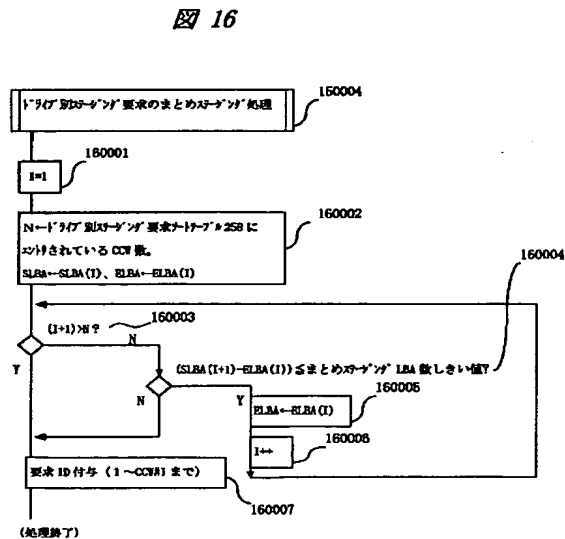
図 12



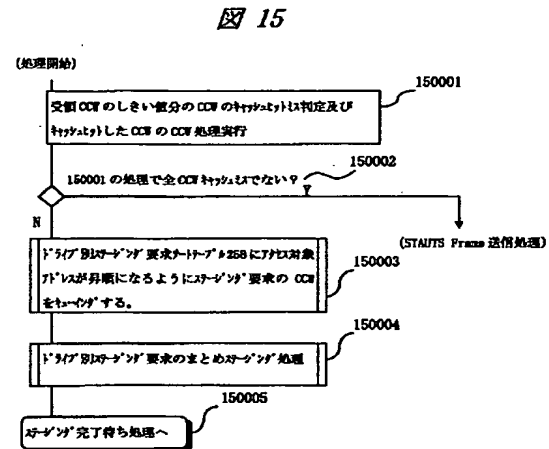
【図13】



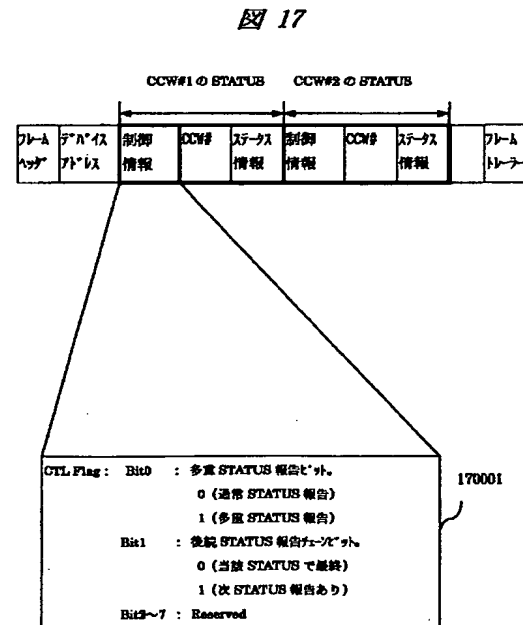
【図16】



【図15】

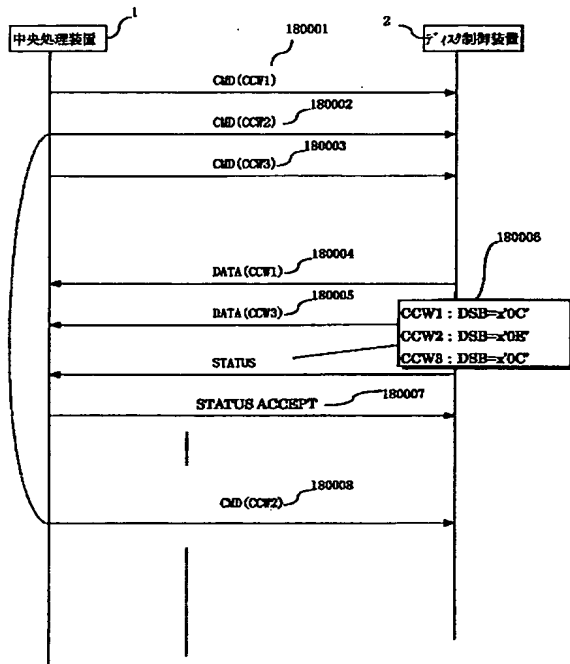


【図17】



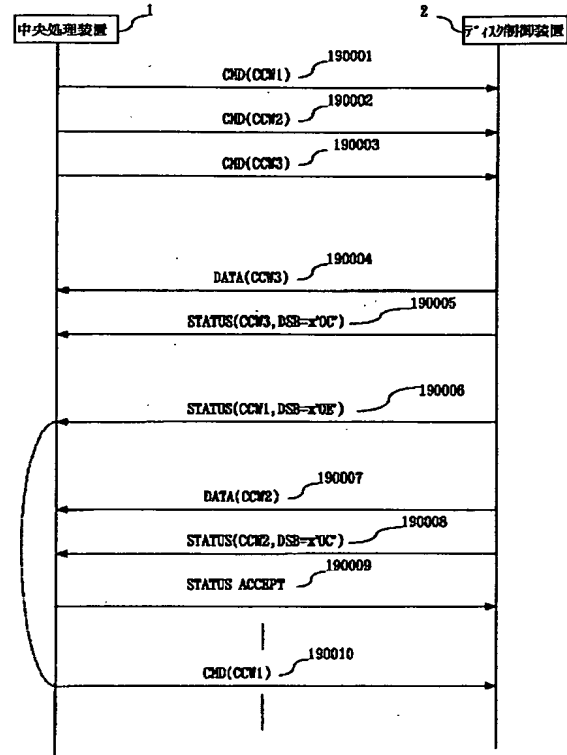
【図18】

図 18



【図19】

図 19



フロントページの続き

(51) Int. Cl.⁷

G 0 6 F 3/06

識別記号

3 0 2

F I

G 0 6 F 3/06

テーマコード (参考)

3 0 2 E